Atty. Dkt. No.: 035905/0134

U.S. PATENT APPLICATION

For

HIGH DENSITY 3D RAIL STACK ARRAYS AND METHOD OF MAKING

Inventors:

Kedar Patel,

Alper Ilkbahar,

Roy Scheuerlein and

Andrew J. Walker

Prepared by:
Foley & Lardner
3000 K St. N.W.
Washington D.C. 20007
(202) 672-5300

HIGH DENSITY 3D RAIL STACK ARRAYS AND METHOD OF MAKING

FIELD OF THE INVENTION

[0001] The present invention is directed generally to semiconductor devices and methods of fabrication and more particularly to three dimensional arrays of thin film transistors and method of fabrication.

BACKGROUND OF THE INVENTION

[0002] Thin film transistors (TFTs) are utilized in various devices, such as a liquid crystal displays, static random access memories (SRAMs) and in nonvolatile memories. Conventional TFTs have a structure that is similar to conventional bulk metal oxide semiconductor field effect transistors (MOSFETs), except that TFTs are formed in a semiconductor layer that is located above an insulating substrate, such as a glass substrate or a semiconductor substrate that is covered by an insulating layer. The TFT device density on the substrate is usually lower than desired. The decreased device density increases the device cost, since fewer devices can be made on each substrate. PCT published application WO 02/15277 A2, which corresponds to U.S. Application Serial No. 09/927,648 filed on August 13, 2002, incorporated herein by reference in its entirety, describes how three dimensional rail stack arrays of TFTs may be used utilized to decrease device density.

BRIEF SUMMARY OF THE INVENTION

[0003] One preferred aspect of the present invention provides a semiconductor device, comprising a first field effect transistor, comprising (i) a first rail comprising a first channel, a first gate insulating layer and a first gate electrode, (ii) a first source region, and (iii) a first drain region.

The device also comprises a second field effect transistor, comprising (i) a second rail comprising a second channel, a second gate insulating layer and a second gate electrode, (ii) a second source region, (iii) a second drain region, wherein the first rail comprises at least one of the second source region or the second drain region.

[0004] Another preferred aspect of the present invention provides a monolithic three dimensional array of field effect transistors, comprising (a substrate and a plurality of first rails disposed at a first height relative to the substrate in a first direction, wherein each of the plurality of first rails comprises a first heavily doped semiconductor layer of a first conductivity type. The array also comprises a plurality of second rails disposed in contact with the first rails, at a second height different from the first height, and in a second direction different from the first direction, wherein each of the plurality of second rails comprises a second heavily doped semiconductor layer of the first conductivity type, and a plurality of third rails disposed in contact with the second rails, in the first direction at a third height relative to the substrate such that the second rails are located between the first and the third rails, wherein each of the plurality of third rails comprises a third heavily doped semiconductor layer of the first conductivity type. Portions of the plurality of second rails comprise gate electrodes of a plurality of first field effect transistors and source or drain regions of a plurality of second field effect transistors.

[0005] Another preferred aspect of the present invention provides a monolithic three dimensional array of field effect transistors, comprising a substrate and a plurality of first rails disposed at a first height relative to the substrate in a first direction, wherein each of the plurality of first rails comprises a first heavily doped semiconductor layer of a first conductivity type. The array also comprises (c) a plurality of second rails disposed at a second height different from the first height, and in a second direction

different from the first direction. Each of the plurality of second rails comprises a second lightly doped semiconductor channel layer of a second conductivity type located in contact with the first rails, a second heavily doped semiconductor layer of the first conductivity type, a second gate insulating layer between the second channel layer and the second heavily doped layer of the first conductivity type, and a second heavily doped semiconductor layer of the second conductivity type electrically connected to the second heavily doped semiconductor layer of the first conductivity type by a metal or a metal silicide layer. The array also comprises a plurality of third rails disposed in the first direction at a third height relative to the substrate. Each of the plurality of third rails comprises a third lightly doped semiconductor channel layer of the first conductivity type located in contact with the second heavily doped layer of the second conductivity type in the second rails, a third heavily doped semiconductor layer of the second conductivity type, a third heavily doped semiconductor layer of the first conductivity type electrically connected to the third heavily doped semiconductor layer of the first conductivity type by a metal or a metal silicide layer, and a third gate insulating layer between the channel layer and the third heavily doped layer of the second conductivity type.

[0006] Another preferred aspect of the present invention provides a semiconductor device, comprising a first field effect transistor of a first polarity and a second field effect transistor of a second polarity. A gate electrode of the first transistor is electrically connected to a source or drain of the second transistor without any lateral interconnects.

[0007] Another preferred aspect of the present invention provides a monolithic three dimensional memory array of field effect transistors, comprising a substrate and a plurality of first rails disposed at a first height relative to the substrate in a first direction, wherein each of the

plurality of first rails comprises a first heavily doped semiconductor layer of a first conductivity type. The array also comprises a plurality of second rails disposed in contact with the first rails at a second height different from the first height, and in a second direction different from the first direction. Each of the plurality of second rails comprises a second heavily doped semiconductor layer of the first conductivity type, a second lightly doped semiconductor channel layer of the second conductivity type, and a second charge storage region located between the second heavily doped semiconductor layer and the second lightly doped semiconductor layer. The array further comprises a plurality of third rails disposed in the first direction at a third height relative to the substrate such that the second rails are located between the first and the third rails. Each of the plurality of third rails comprises a third heavily doped semiconductor layer of the first conductivity type, a third lightly doped semiconductor channel layer of the second conductivity type, and a third charge storage region located between the third heavily doped semiconductor layer and the third lightly doped semiconductor layer. The second lightly doped semiconductor layers in the second rails contact the first heavily doped semiconductor layers in the first rails. The third lightly doped semiconductor layers in the third rails contact the second heavily doped semiconductor layers in the second rails.

[0008] Another preferred aspect of the present invention provides a method of making a monolithic three dimensional field effect transistor array, comprising forming a plurality of first rails disposed at a first height relative to a substrate in a first direction, wherein each of the plurality of first rails comprises a first heavily doped semiconductor layer of a first conductivity type, forming a first insulating isolation layer over the first plurality of rails and patterning the first isolation layer to form a plurality of first openings exposing upper portions of first rails. The method

further comprises forming a second lightly doped semiconductor layer of a second conductivity type over the patterned isolation layer such that transistor channel portions in the second lightly doped layer of the second conductivity type contact the first heavily doped layer of the first conductivity type through the first openings. The method further comprises forming a second gate insulating layer over the second lightly doped semiconductor layer of the second conductivity type, forming a second heavily doped semiconductor layer of the first conductivity type over the gate insulating layer, and patterning the second heavily doped layer of the first conductivity type, the second gate insulating layer and the second lightly doped layer of the second conductivity type to form a plurality of second rails extending in a second direction different from the first direction.

[0009] Another preferred aspect of the present invention provides a monolithic three-dimensional array of active devices comprising odd and even levels of field effect transistors, wherein odd levels comprise transistors of a first polarity, even levels comprise transistors of a second polarity and each transistor comprises a gate electrode, source, and drain, wherein the gate electrodes, sources, and drains of the transistors of at least two levels comprise polysilicon. Current flows between the source and the drain in a first direction through transistors of the first polarity and current flows between the source and the drain in a second direction not parallel to the first direction through transistors of the second polarity.

[0010] Another preferred aspect of the present invention provides a semiconductor device, comprising a first transistor having a gate electrode, source, channel, and drain oriented in a first direction, and a second transistor having a gate electrode, source, channel, and drain oriented in a second direction different from said first direction. The gate

electrode of said first transistor and the source of said second transistor are disposed in a portion of a first rail.

[0011] Another preferred aspect of the present invention provides a semiconductor device a semiconductor device comprising a first rail, the first rail comprising a gate electrode of a first field effect transistor, and a source or drain of a second field effect transistor. The first transistor and the second transistor are oriented in non-parallel directions.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Figure 1 is a three dimensional view of an array of the first preferred embodiment of the present invention.

[0013] Figures 2-3 are three dimensional views of portions of the array of the first preferred embodiment of the present invention.

[0014] Figures 4-7 are schematic diagrams illustrating how various circuit elements can be made using the array of the first preferred embodiment.

[0015] Figures 8 and 9 are three dimensional views of an array of the second preferred embodiment of the present invention.

[0016] Figures 10A-D are side cross sectional views of steps in a method of making of an array of the preferred embodiments of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0017] The present inventors have realized that the device density may be further increased in a rail stack array of TFTs if one rail is contains a gate of one TFT and a source/drain of another TFT.

[0018] **THE ARRAY OF THE FIRST PREFERRED EMBODIMENT.** Figure 1 illustrates a monolithic three dimensional array 1 of field effect transistors according to the first preferred embodiment. The array 1 is formed over a substrate (not shown for clarity in Figure 1). The array contains a plurality of first rails 3 disposed at a first height relative to the substrate in a first direction. For example, the first rails 3 are illustrated in Figure 1 as being at the lowest height above the substrate and extend diagonally to the left into the page. Of course, any other suitable height and direction may be used instead. Each of the first plurality of rails 3 comprises at least a first heavily doped semiconductor layer of a first conductivity type. For example, if the first conductivity type is n-type, then each first rail 3 comprises an N+ polysilicon layer 5. Of course, the first conductivity type may be p-type, if desired. Preferably, another optional N+ polysilicon layer 7 is located under layer 5 in the first rails 3 and an optional metal or a metal silicide layer 9 is located between layers 5 and 7 to increase the conductivity of the first rails 3.

[0019] The array 1 further comprises a plurality of second rails 13 disposed in contact with the first rails 3, at a second height different from the first height. For example, rails 13 may be located directly above rails 3. The second rails 13 are disposed in a second direction different from the first direction. For example, the second rails 13 are illustrated in Figure 1 extending diagonally to the right into the page. Thus, rails 13 are disposed substantially perpendicular to rails 3, such as at 70-110 degrees, preferably at 80-100 degrees, most preferably at 90 degrees with respect to each other. Of course, any other suitable height and direction may be used instead. Each of the second plurality of rails 13 comprises a second heavily doped semiconductor layer of the first conductivity type 15, such as an N+ polysilicon layer.

[0020] The array 1 further comprises a plurality of third rails 23 disposed in contact with the second rails 13, in the first direction (i.e., parallel to the first rails 3 and perpendicular to the second rails 13). However, the third rails 23 may be disposed in another direction that is not parallel to the first direction of the first rails 3. Rails 23 are disposed at a third height relative to the substrate such that the second rails 13 are located between the first 3 and the third rails 13. Each of the third plurality of rails 23 comprises a third heavily doped semiconductor layer of the first conductivity type 25, such as an N+ polysilicon layer.

[0021] In the array 1 of the first preferred embodiment, the second rails 13 also contain a second lightly doped semiconductor channel layer 16 of a second conductivity type, such as P- polysilicon layer. Layer 16 is disposed in contact with the first rails 3, such as in contact with N+ layers 5 in the first rails 3. The second rails 13 also contain a gate insulating layer 17 between the channel layer 16 and the second heavily doped layer of the first conductivity type 15. The second rails 13 further contain a second heavily doped semiconductor layer of the second conductivity type 18. For example, layer 18 may comprise a P+ polysilicon layer. Layer 18 is electrically connected to the second heavily doped semiconductor layer of the first conductivity type 15 by a metal or a metal silicide layer 19.

[0022] In the array 1 of the first preferred embodiment, the plurality of third rails 23 also contain a third lightly doped semiconductor channel layer of the first conductivity type 26, such as an N- polysilicon layer. Layer 26 is disposed in contact with the second heavily doped layer of the second conductivity type 18 in the second rails 13. The third rails 23 also contain a third heavily doped semiconductor layer of the second conductivity type 28. Layer 28 may be a P+ polysilicon layer, for example. Layer 28 is electrically connected to the third heavily doped

semiconductor layer of the first conductivity type 25 by a metal or a metal silicide layer 29. A gate insulating layer 27 is disposed between the channel layer 26 and the third heavily doped layer of the second conductivity type 28. Furthermore, a planarized insulating fill layer 30 (shown as white space in Figure 1) is located between adjacent first, second and third rails.

[0023] The rails 3, 13 and 23 are illustrated as having a rectangular cross section. However, the rails 3, 13 and 23 may have any other desired cross section, such as square, triangular, polygonal, oval and/or circular cross section. The cross section may be constant or variable along the length of the rail.

[0024] The rails 3, 13, 23 contain a plurality of field effect transistors, such as TFTs at the intersections of the rails. Thus, portions of the plurality of second rails 13 comprise gate electrodes of a plurality of first field effect transistors 31 and source or drain regions of a plurality of second field effect transistors 33.

[0025] In the array 1 of the first embodiment, the second heavily doped semiconductor layer of the first conductivity type 15 comprises a gate electrode of the first transistors 31. The second heavily doped semiconductor layer of the second conductivity type 18 comprises a source or drain region of the second transistors 33. The third heavily doped semiconductor layer of the second conductivity type 28 in the third rails 23 comprises a gate electrode of the second transistors 33. The first heavily doped semiconductor layer of the first conductivity type 5 in the first rails 3 comprises a source or drain region of the first transistors 31.

[0026] The first 3 and the second 13 rails comprise a first transistor level 35 formed in or above the substrate containing the plurality of first transistors 31. The second 13 and the third 23 rails comprise a second

transistor level 37 located above the first level 35 containing the plurality of second transistors 33. Thus, the first 35 and the second 37 levels overlap, and both include the second rails 13.

[0027] Each rail 13, 23 contains four layers, including a channel layer, a gate insulating layer and two heavily doped semiconductor layers of opposite conductivity type. The conductivity type of the channel layer switches in each successive rail. Furthermore, the order of the two heavily doped semiconductor layers of opposite conductivity type is reversed in each successive rail. Each transistor level 35, 37 contains TFTs of one conductivity type. The conductivity type of the TFTs is reversed in each succeeding transistor level.

[0028] The array 1 is not limited to three sets of rails and two transistor levels. One or more additional sets of rails and transistor levels may be included in the array 1. For example, the array 1 may also include a plurality of fourth rails disposed in the second direction. Thus, the fourth rails are disposed parallel to the second rails and perpendicular to the first and third rails. The fourth rails are disposed at a fourth height relative to the substrate such that the third rails 23 are located between the second 13 and the fourth rails. The fourth rails are the same as the second rails 13 and are omitted from Figure 1 for clarity.

[0029] The third and the fourth rails comprise a third transistor level located above the second level containing the plurality of third transistors. Thus, portions of the plurality of third rails 23 comprise gate electrodes of the second field effect transistors 33 and source or a drain regions of a plurality of third field effect transistors (not shown for clarity). Portions of the fourth rails comprise gate electrodes of the third field effect transistors. Additional rails and transistor levels may be provided in the array as desired, in the same manner as the first four rails. For example,

the array 1 may contain three to nine rails and two to eight transistor levels.

[0030] Only five rails are needed to form two transistors. Rail one contains the gate and channel of a first transistor and source or drain of a second transistor. Rail two contains the gate and channel of the second transistor. Rail three contains the drain or source of the second transistor. Rails four and five contain the source and drain of the first transistor. This increases the density of the TFTs, since more TFTs may be packed into a given space above the substrate.

[0031] Details of one first transistor (i.e., n-type metal oxide semiconductor (NMOS) TFT) 31 and one second transistor (i.e., p-type metal oxide semiconductor (PMOS) TFT) 33 are illustrated in Figure 2. The first field effect transistor 31 comprises a first gate electrode 41 comprising a portion of one of the second rails 13. Specifically, gate electrode 41 comprises a portion of the N+ polysilicon layer 15. Transistor 31 further contains a first channel region 43 comprising a portion of the first lightly doped semiconductor layer 16 (P- polysilicon layer) located in the same one of the second rails 13 and a portion of the first gate insulating layer 17 as the transistor gate insulating layer. A first source region 45 comprises a portion of the N+ polysilicon layer 5 of one of the first rails 3. A first drain region 47 comprises a portion of the N+ polysilicon layer 5 of another one of the first rails 3. Thus, TFT 31 comprises an NMOS TFT.

[0032] The second field effect transistor 33 comprises a second gate electrode 51 comprising a portion of one of the third rails 23. Specifically, gate electrode 51 comprises a portion of the P+ polysilicon layer 28. Transistor 33 further contains a first channel region 53 comprising a portion of the second lightly doped semiconductor layer (N-

polysilicon layer) 26 located in the same one of the third rails 23 and a portion of the second gate insulating layer 27 as the transistor gate insulating layer. A second source region 55 comprises a portion of the P+ polysilicon layer 18 in one of the second rails 13. A second drain region 57 comprises a portion of the P+ polysilicon layer 18 in another one of the second rails 13. Thus, transistor 33 is a PMOS TFT. The first 31 and the second 33 transistors comprise top gate staggered TFTs located above an insulating substrate or above an insulating layer formed over a silicon substrate.

[0033] The array 1 also contains insulating isolation layers which isolate adjacent transistors from each other, as illustrated in Figure 2. A first insulating isolation layer 48 is located between the first 3 and the second 13 rails. A second insulating isolation layer 58 is located between the second 13 and the third 23 rails. A plurality of first openings 49 are located in the first isolation layer 48. The second lightly doped semiconductor layer of the second conductivity type 16 (P- polysilicon layer) in the second rails 13 contacts the first heavily doped semiconductor layer of the first conductivity type 5 (N+ polysilicon layer) in the first rails 3 through openings 49. A plurality of second openings 59 are located in the second isolation layer 58. The third lightly doped semiconductor layer of the first conductivity type 26 (N- polysilicon layer) in the third rails 23 contacts the second heavily doped semiconductor layer of the second conductivity type 28 in the second rails 13 through openings 59. Thus, the channel regions of the TFTs are formed in the openings 58, 59, such that the respective insulating isolation layers are adjacent to lateral edges of the respective channel regions to form island channel regions.

[0034] Figure 3 illustrates how individual transistors in array 1 are interconnected to form a logic or memory device. A plurality of vias 60,

61 extend through the first 17 and second 27 gate insulating layers and the first 16 and the second 26 channel layers, respectively. The respective heavily doped layers 15, 28 contact the respective heavily doped layers of the same conductivity type 5, 18 in the rails located below through the vias 60, 61. For example, a portion of layer 28 in via 61 acts as an interconnect between the gate 51 and a source 55 or drain 57 of a second transistor 33. This interconnect electrically connects the second and the third rails. Similar vias are formed in other rails to form the desired interconnection.

[0035] Figures 4 through 6 illustrate some of the circuit elements that may be formed by connecting heavily doped semiconductor layers in adjacent rails through the vias. Figure 4A illustrates a circuit schematic of two transistor device where a gate electrode 28 of a first transistor 33 of a first polarity (i.e., the PMOS gate electrode) is electrically connected to a source or drain 25 of a second transistor 31 of a second polarity (i.e., the source or drain of the NMOS) without any lateral interconnects. Figure 4B illustrates a device implementation of the circuit schematic of Figure 4A. For example, rails 13 in Figure 4B correspond to the second rails 13 in Figure 1 and rails 23 correspond to the third rails 23 in Figure 1. Rails 23 overly rails 13. Rail(s) 73 overly rails 23. The elements in rail(s) 73 are the same as in rails 13. Rail 73 contains an N+ polysilicon layer 75, a P- polysilicon channel layer 76, a gate insulating layer 77, a P+ polysilicon layer and a metal or metal silicide layer 79.

[0036] The P+ layer 28 in rail 23 acts as a gate electrode of the PMOS transistor 33, while the N+ layer 25 in the same rail 23 acts as a source or drain of the NMOS transistor 31. Since the gate 28 of the PMOS transistor 33 and the source or drain 25 of the NMOS transistor 31 are located in the same rail 23, the PMOS gate 28 is electrically connected to the NMOS source or drain 25 using metal or metal silicide

layer 29 without any lateral interconnects. The N+ layer 75 in rail 73 acts as the gate of the NMOS transistor 31, while the P+ layers 18 in rail 13 act as source and drain regions of the PMOS transistor 33.

[0037] Figure 5A illustrates a circuit schematic of two transistor inverter, such as a complementary metal oxide semiconductor (CMOS) inverter. Figure 5B illustrates a device implementation of the circuit schematic of Figure 5A. For example, rails 13 in Figure 5B correspond to the second rails 13 in Figure 1 and rails 23 correspond to the third rails 23 in Figure 1. Rail(s) 73 are the same as that shown in Figure 4B. Rails 23 overly rails 13. Rail(s) 73 overly rails 23.

[0038] The inverter contains an input 63 into the NMOS 31 and PMOS 33 gate electrodes and an output 65 from the drains of the NMOS 31 and PMOS 33. The source of the PMOS is connected to voltage V_{DD} 67 while the source of the NMOS is connected to ground 69. A first via 60 between rails 13 and 23 allows an electrical connection to be made between the P+ layer 18 in rail 13 acting as a drain of PMOS 33 and the N+ layer 25 in rail 23 acting as a drain of the NMOS 31. A second via 60 between rails 23 and 73 allows an electrical connection to be made between the P+ layer 28 in rail 23 acting as a gate of PMOS 33 and the N+ layer 75 in rail 73 acting as a gate of the NMOS 31. The P+ layer 18 in another rail 13 which acts as a source of the PMOS 33 is connected to voltage V_{DD} 67, while the N+ layer 25 in another rail 23 which acts as the source of the NMOS is connected to ground 69. In the inverter, the gate of the PMOS 33 is located in rail 23 which does not contain a source or drain of the NMOS 31. Thus, three rails 23 are used to form a two transistor inverter.

[0039] Figure 6A illustrates a circuit schematic of a six transistor CMOS static random access memory (CMOS SRAM). Figure 6B

illustrates a device implementation of the circuit schematic of Figure 6A. For example, rails 13 in Figure 6B correspond to the second rails 13 in Figure 1 and rails 23 correspond to the third rail 23 in Figure 1. Rail(s) 73 are the same as rails 23. Rails 13 overly rails 73. Rails 23 overly rails 13. The gate electrodes of the two NMOS access transistors 31 are connected to the common word line WL, while their drains are connected to bit lines BL and BLB. The drains of the load PMOS transistors 33 are connected to voltage V_{DD} while the drains of the NMOS driver transistors 31 are connected to ground GND. Vias 60 between rails 73 and 13 and vias 61 between rails 13 and 23 provide the connections to form the SRAM.

[0040] The rails 3, 13, 23, 73 have been illustrated in Figures 1 and 4-6 as extending in a single direction. However, the direction of rails may change over different regions of the substrate, as illustrated in Figure 7. For example, the rails may extend in one direction over one portion of the substrate, and then extend in another direction over a different portion of the substrate. Thus, the transistors located in the same transistor level may be oriented in different directions if desired.

[0041] Furthermore, the rails are preferably continuous and have no gaps or discontinuities. However, if desired, the rails may be discontinuous and may contain one or more gaps or discontinuities, as shown in Figure 7. The rails have also been illustrated as having a rectangular cross section. However, if desired, the rails may have any other desired polygonal, oval or circular cross section. Furthermore, while the rails preferably have the same cross sectional dimensions and shape along their entire length, these dimensions and shape may vary along the length of the rails if desired.

[0042] **THE ARRAY OF THE SECOND PREFERRED EMBODIMENT.** Figure 8 illustrates a three dimensional array 100 of devices according to the second preferred embodiment of the present invention. The array 100 differs from array 1 in that the heavily doped semiconductor layer of the second conductivity type is omitted from each rail. Thus, each transistor level contains transistors of the same conductivity type. In contrast, the array 1 contains transistors of different conductivity type. Furthermore, in the array 100, the heavily doped semiconductor layer of the same conductivity type in each rail serves as a gate of one transistor and a source or drain of another transistor. In contrast, in array 1, different heavily doped semiconductor layers of opposite conductivity type in the same rail serve as a gate of one transistor and a source or drain of another transistor. Therefore, array 1 is preferably used as an array of logic devices, while array 100 is preferably uses as an array of memory devices, such as an array of charge storage devices (i.e., programmable read only memory (PROM), electrically programmable read only memory (EPROM) or electrically erasable programmable read only memory EEPROM). In the case of EPROM and EEPROM, each transistor in the array 100 contains a charge storage region. However, if desired, the array 1 may also be used in an EPROM or EEPROM.

[0043] Thus, in the array 100, a plurality of first rails 103 are disposed at a first height relative to the substrate in a first direction. Each of the first plurality of rails 103 comprises a first heavily doped semiconductor layer of a first conductivity type 105, such as an N+ polysilicon layer. Preferably, another optional N+ polysilicon layer 107 is located under layer 105 in the first rails 103 and an optional metal or a metal silicide layer 109 is located between layers 105 and 107 to increase the conductivity of the first rails 103.

[0044] A plurality of second rails 113 are disposed in contact with the first rails 103 at a second height different from the first height, and in a second direction different from the first direction. Each of the second plurality of rails 113 comprises a second heavily doped semiconductor layer of the first conductivity type 115, such as an N+ polysilicon layer and a second lightly doped semiconductor channel layer of the second conductivity type 116, such as a P- polysilicon layer.

[0045] Each second rail also contains a second gate insulating layer. Preferably, the array 100 contains charge storage transistors, and the second gate insulating layer comprises a portion of second charge storage region 117 located between and in contact with the second heavily doped semiconductor layer 115 and the second lightly doped semiconductor layer 116. The second charge storage region 117 comprises one of a dielectric isolated floating gate, a silicon oxide/silicon nitride/silicon oxide (ONO) dielectric film, an insulating layer containing conductive nanocrystals or any other desired charge storage material.

[0046] The array 100 also contains a plurality of third rails 123 disposed in the first direction at a third height relative to the substrate, such that the second rails 113 are located between the first 103 and the third 123 rails. Each of the third plurality of rails 123 comprises a third heavily doped semiconductor layer of the first conductivity type 125, such as an N+ polysilicon layer and a third lightly doped semiconductor channel layer of the second conductivity type 126, such as a P- polysilicon layer. The third rails 123 also comprise a third gate insulating layer. Preferably, the array 100 contains charge storage transistors, and the third gate insulating layer comprises a portion of a third charge storage region 127 located between and in contact with the third heavily doped semiconductor layer 125 and the third lightly doped semiconductor layer 126. Charge storage region 127 may comprise one of a dielectric

isolated floating gate, an ONO dielectric film, an insulating layer containing conductive nanocrystals or any other desired charge storage material.

[0047] The second lightly doped semiconductor layer 116 in the second rails 113 contacts the first heavily doped semiconductor layer 105 in the first rails 103. The third lightly doped semiconductor layer 116 in the third rails 123 contacts the second heavily doped semiconductor layer 125 in the second rails 123.

[0048] A plurality of first 131 and second 133 transistors are formed above each other in array 100 similar to the first array 1. Each rail except the top and the bottom rail is part of two transistor levels.

[0049] Similar to the array 1 of the first preferred embodiment, the array 100 of the second preferred embodiment contains a first insulating isolation layer 148 located between the first rail 103 and the second rail 113. The first insulating isolation layer 148 is located in the second rails below a first charge storage region 117 adjacent to lateral edges of the first channel regions to form an island first channel regions 161. A plurality of first openings 149 are located in the first insulating isolation layer 148, with the island channel regions 161 being disposed in the openings 149.

[0050] A second insulating isolation layer 158 is located between the second rail 113 and the third rail 123. The second insulating isolation layer 158 is located in the third rails below the second charge storage region 127 adjacent to lateral edges of a second channel regions to form island second channel regions 163. A plurality of second openings 159 are located in the second insulating isolation layer 158, with the island channel regions 163 being disposed in the openings 159.

[0051] A plurality of first transistor 131 island channels 161 comprise portions of the second lightly doped semiconductor layer 116 located in the first openings 149. The first transistor island channels 161 contact the first heavily doped semiconductor layer 105 in the first rails 103. First transistor bit lines comprise the first heavily doped semiconductor layer 105 in the first rails 103. First transistor word lines comprise the second heavily doped semiconductor layer 115 in the second rails 113.

[0052] A plurality of second transistor 133 island channels 163 comprise portions of the third lightly doped semiconductor layer 126 located in the second openings 159. The second transistor island channels 163 contact the second heavily doped semiconductor layer 115 in the second rails 113. The second transistor bit lines comprise the second heavily doped semiconductor layer 115 in the second rails 113. The second transistor word lines comprise third heavily doped semiconductor layers 125 in the third rails 123.

[0053] In addition, the array 100 contains a plurality of fourth rails 173 disposed in contact with the third rails 123. The fourth rails 173 are disposed at a fourth height above the third rails 123 and extend in the second direction, parallel to the second rails 113. Each of the plurality of fourth rails comprises a fourth heavily doped semiconductor layer of the first conductivity type 175, a fourth lightly doped semiconductor layer of the second conductivity type 176 and a fourth charge storage region 177 located between and in contact with layers 175 and 176. As with the array 1 of the first embodiment, a third plurality of TFTs 135 are located in a third transistor level encompassing the third 123 and fourth 173 rails.

[0054] A planarized insulating fill layer 130 is located between adjacent rails. If desired, an optional metal or metal silicide layer 109,

119, 129 may be provided inside the heavily doped semiconductor layers 105, 115 and 125. In this case, the heavily doped semiconductor layers 105, 115 and 125 comprise upper and lower sublayers separated by the metal or metal silicide layer.

[0055] Using the architecture of Figure 8, the memory density is improved to an effective cell size of $8f^2/(n-1)$, where n is the number of semiconductor layers. The dual purpose of heavily doped semiconductor layers as gates and source/drains provides for a different implementation of the erase and read/write steps. Traditional row and column circuits are optimized to perform single operations, whereas the architecture illustrated in Figure 8 provides that row and column operations to both erase as well as read/write.

[0056] **THE METHOD OF BIASING THE ARRAY OF THE SECOND PREFERRED EMBODIMENT TO WRITE, READ AND ERASE DATA.** The method of biasing the memory array 100 shown in Figure 8 to make read, write and erase operations will now be explained with reference to Figure 9. While absolute biases across transistors will be illustrated for clarity of explanation, the biasing of the array may be implemented as a sum of positive and negative potentials with respect to ground. The primary mechanism of write and erase operations in the array 100 of is carrier tunneling, such as Fowler-Nordheim or modified Fowler-Nordheim tunneling, where the injection current has a near-exponential dependence on the applied voltage. Hence, an inhibit voltage, which is lower than the write or erase voltage, can be applied to non-selected transistors in array 100 without significantly disturbing the stored value in these transistors.

[0057] Figure 9 illustrates the naming convention of the voltages applied to the array 100 when writing, reading or erasing data to and from the selected transistor 131 (circled in Figure 9). Specifically, VGsel

is the gate voltage for the selected transistor. VBsel is the bit line voltage for the selected transistor. VGunsel is gate voltage for the unselected transistors. VBunsel is the bit line voltage for the unselected transistors.

[0058] The write operation to the targeted or selected transistor 131 having its gate in one third rail 123 and its source and drain in adjacent second rails 113 in array 100 is as follows. A high programming voltage, VPP, is provided to the gate of the targeted transistor 131 with its drain and source grounded. All other third rails 123 are biased at Vinhbtg, the gate inhibit voltage. This prevents unintentional programming of the cells sharing the same bit lines with the programmed transistor 131. VPP is a voltage across the gate that will program a cell, while Vinhbtg is a voltage that will not. All bit lines in rails 113 except for the rails which act as the source and drain of the programmed transistor 131 are maintained at Vinhbtb, the bit line inhibit voltage. Vinhbtb is a voltage applied to the bit lines of a magnitude such that applying VPP to the selected word line will not program transistors underneath this word line. Thus, voltages Vinhbtg and Vinhbtb are lower than voltage VPP. All first rails 103 and fourth rails 173 are biased Vinhbtb.

[0059] The exemplary voltages applied to the array 100 for a write operation to the selected transistor 131 are illustrated in Table 1 below. Of course other specific voltages values may be used instead, as long as the relationship of the different voltages (i.e., ground, float, low and high) remain the same.

Table 1

VGsel	VPP (10V)
VBsel	Ground (0V)
VGunsel	Vinhbtg (5V)
VBunsel	Vinhbtb (5V)

[0060] The erase operation of the targeted or selected transistor 131 having its gate in one third rail 123 and its source and drain in adjacent second rails 113 in array 100 is as follows. The gate of the targeted transistor 131 in rail 123 is grounded (zero volts are applied to the gate) and a high erase voltage, VEE, is applied to the source and drain located in adjacent second rails 113 of targeted transistor 131. All other third rails 123 are biased at Vinhbtg to inhibit erase of other transistors at the same level as the targeted transistor 131. All other second rails 113 are biased at Vinhbtb to prevent erasing of the non-targeted transistors. VEE is sufficient to erase a transistor, while Vinhbtb is not. Thus, voltage VEE is higher than voltages Vinhbtb and Vinhbtg. The above described bias condition is suitable for erasing a single transistor (i.e., cell or bit). To erase the contents of an entire block of transistors, voltage VEE is applied to all bit lines of the selected block of transistors. This will erase all the cells controlled by the select gate simultaneously. Depending on the selected scheme, lines in other rails can be either grounded or left at Vinhbtb.

[0061] The exemplary voltages applied to the array 100 for an erase operation to the selected transistor 131 are illustrated in Table 2 below. Of course other specific voltages values may be used instead, as long as the relationship of the different voltages (i.e., zero or float, low and high) remain the same.

Table 2

VGsel	Ground (0V)
VBsel	VEE (10V)
VGunsel	Vinhbtg (5V)
VBunsel	Vinhbtb (5V)

[0062] The read operation of the targeted or selected transistor 131 having its gate in one third rail 123 and its source and drain in adjacent second rails 113 in array 100 is as follows. The read operation is performed by sensing the current between the drain and source terminals of the targeted transistor 131. The control gate of the targeted transistor 131 is raised to the appropriate voltage, VRDG, while biasing the drain and source of the transistor to VRDD and VRDS, respectively. The exact values of these voltages depend on the device characteristics and their statistical distribution. For example, voltage VRDG may be about 2-3V, voltage VRDD may be about 2V and voltage VRDS may be about 0.5V. Voltages VRDG and VRDD are higher than voltage VRDS. Voltages VRDG, VRDD and VRDS are lower than the voltages required to program and erase the transistor. All other rails are left to float.

[0063] The exemplary voltages applied to the array 100 for a read operation to the selected transistor 131 are illustrated in Table 3 below. Of course other specific voltages values may be used instead, as long as the relationship of the different voltages (i.e., zero or float, low and high) remain the same.

Table 3

VGsel	VRDG
VBsel	VRDD/VRDS
VGunsel	Float
VBunsel	Float

[0064] Thus, in summary, a predetermined transistor of the memory array is programmed by applying a high programming voltage to the transistor's word line, grounding the transistor's bit lines and applying a low programming inhibiting voltage to unselected word lines and bit lines. This transistor is erased by applying a high erase voltage to the

transistor's bit lines, grounding the transistor's word line and applying a low erase inhibiting voltage to the word lines in the same transistor level as the predetermined transistor's word line and to the bit lines in the same transistor level as the predetermined transistor's bit lines. This transistor is read by applying a first low read voltage to the transistor's word line and drain bit line, applying a second read voltage which is lower than the first read voltage to the transistor's source bit line, allowing the word lines in the same transistor level as the predetermined transistor's word line and the bit lines in the same transistor level as the predetermined transistor's bit lines to float, and sensing a current between the predetermined transistor's bit lines.

[0065] **PREFERRED FEATURES OF THE ARRAYS OF THE FIRST AND SECOND EMBODIMENTS.** Preferably, the arrays 1 and 100 comprise monolithic three dimensional arrays of devices. The term "monolithic" means that layers of each level of the array were directly deposited on the layers of each underlying level of the array. Thus, the first rails 3, 103 are monolithically located above the substrate, the second rails 13, 113 are monolithically located on the first rails 3, 103, the third rails 23, 123 are monolithically located on the second rails 13, 113, and the fourth rails 173 are monolithically located on the third rails 123. Less preferably, two dimensional arrays may be formed separately and then packaged together to form a three dimensional non-monolithic device array.

[0066] As shown in Figures 1 and 8, odd transistor levels comprise transistors 31, 131 of a first polarity (i.e., level 35 shown in Figure 1 comprises NMOS transistors), while even transistor levels comprise transistors 33, 133 of a second polarity (i.e., level 37 comprises PMOS transistors). Some or all transistors in different levels are oriented in different directions. Thus, current flows between the source and the

drain in a first direction through transistors 31 of the first polarity, while current flows between the source and the drain in a second direction not parallel to the first direction through transistors 33 of the second polarity. Preferably, the second direction is substantially orthogonal to the first direction. Thus, the first and the second transistors are disposed in different directions even when the gate electrode of the first transistor and the source or drain of the second transistor are disposed in a portion of the same rail.

[0067] The arrays 1 and 100 were illustrated as having top gate staggered TFTs. However, if desired, the arrays 1 and 100 may be formed to contain bottom gate staggered TFTs. In this case, arrays 1 and 100 are formed upside down (i.e., the order of formation of the rails is reversed) over the substrate. Of course, a three dimensional monolithic array may comprise both top gate and bottom gate TFTs, such as having one level of the array containing top gate TFTs and another level of the array containing bottom gate TFTs.

[0068] The arrays 1, 100 were illustrated as comprising TFTs. However, if desired, the first rails 5 may be formed in a single crystal silicon substrate, such that the first transistors 31 comprise bulk silicon MOSFETs. Furthermore, arrays 1 and 100 may be incorporated into the same device if desired. Thus, array 1 may be formed above, below or adjacent to array 100 on the same substrate. If desired, the conductivity types of each semiconductor layer may be reversed (i.e., p and n type switched).

[0069] Preferably, the semiconductor layers described above comprise polysilicon, but may comprise amorphous silicon or other semiconductor materials if desired. Preferably, the gate insulating layers and isolation layers comprise an insulating layer, such as silicon dioxide,

silicon oxynitride, silicon nitride or aluminum oxide. Preferably, the planarized insulating fill layer 30, 130 comprises a silicon dioxide, silicon oxynitride, silicon nitride, spin-on glass, borophosphosilicate glass (BPSG), PSG or BSG layer. The metal layers 9, 19 and 29 may comprise aluminum, copper, tungsten or titanium (including titanium nitride). The metal silicide layers 9, 19 and 29 may comprise any silicide, such as titanium, tungsten, cobalt, platinum or nickel silicide. The substrate may comprise a semiconductor substrate, such as a monocrystalline silicon or a gallium arsenide substrate or an insulating substrate, such as a glass, quartz, plastic or ceramic substrate. If desired, an insulating layer, such as silicon dioxide, silicon oxynitride, silicon nitride or aluminum oxide, may be formed over the substrate.

[0070] The arrays 1, 100 may be used in any device or system, such as a liquid crystal display (either in the driver or in the active matrix portion), in a logic device or in a memory device, such as an SRAM, a dynamic random access memory (DRAM) or a nonvolatile read only memory ("ROM"), such as a PROM (i.e., mask ROM), EPROM or EEPROM.

[0071] **METHOD OF MAKING THE ARRAYS OF THE FIRST AND SECOND EMBODIMENTS**. The arrays 1 and 100 may be made by any desired method. A preferred method of making monolithic three dimensional arrays 1 and 100 is illustrated in Figures 10A-D. All layers may be deposited by any desired method, such as chemical vapor deposition, sputtering, molecular beam epitaxy, etc. All patterning steps may be carried out by photolithography and wet or dry etching. Element numbers in Figures 10A-D refer to the elements in array 1 of Figure 1 for explanation. However, the method of Figures 10A-D may be used to form the array 100 shown in Figure 8 instead, as will be described below.

[0072] A plurality of first rails 3 are formed in or over a substrate 2. For example, the first rails 3 may be formed by depositing a first N+ or P+ polysilicon layer 7, a metal or metal silicide layer 9, such as titanium or cobalt and a second N+ or P+ polysilicon layer 5, in that order, over a substrate 2. Layers 5 and 7 may be in-situ doped during deposition or doped after deposition by ion implantation. A photoresist mask is formed over layer 5 and layers 5, 7 and 9 are then etched to form the first rails 3, as shown in Figure 10A. If desired, the rails 3 may be annealed at any time in the process to react the metal layer 9 with the adjacent polysilicon layers to form a metal silicide layer.

[0073] A first insulating fill layer 30 is then deposited over and between the first rails 3. Layer 30 is then polished by chemical mechanical polishing using the first rails as a polish stop to expose the first rails 3. Layer 30 remains between the rails 3. Alternatively, layer 30 may be planarized with the top portions of the first rails 3 using etchback instead of polishing.

[0074] A first insulating isolation layer 48 is then deposited over the first plurality of rails 3 and over the planarized fill layer 30. The first isolation layer 48 is then patterned by photolithography and etching to form a plurality of first openings 49 exposing upper portions of adjacent first rails 3, as illustrated in Figure 10B.

[0075] Alternatively, deposition of a separate isolation layer 48 may be omitted if desired. Instead, the upper portion of the fill layer 30 may be left over the rails 3 to function as the isolation layer. The openings 49 are then formed in the upper portion of the fill layer 30 to expose the first rails 3.

[0076] A second lightly doped semiconductor layer 16 of a opposite conductivity type to layer 5 is formed over the patterned isolation layer

48, as illustrated in Figure 10C. For example, layer 16 comprises P-polysilicon or amorphous silicon layer if layer 5 comprises an N+ polysilicon layer. Layer 16 may be doped in-situ or by ion implantation. If desired, layer 16 may be annealed at any time during the process to recrystallize the amorphous silicon into polysilicon or to increase the grain size of the as-deposited polysilicon layer. Annealing may also be optionally used to outdiffuse dopants from layer 5 into layer 16 to form source and drain regions which extend from layer 5 into layer 16. Layer 16 may be annealed in a furnace or by using laser or flash lamp irradiation. If desired, an optional metal or metal silicide catalyst material may be used to increase the polysilicon grain size and to allow the use of a lower annealing temperature.

[0077] Portions of layer 16 located in openings 49 comprise transistor channel. As shown in Figure 10C, layer 16 extends over layer 48 without interruption. However, if desired, layer 16 may be optionally etched or polished to expose the top portion of layer 48, such that portions of layer 16 are located only in the openings 49. This polishing or etchback step allows formation of discrete channel islands. The discrete channel islands 161 are preferred in the array 100 of the second embodiment. In another alternative method of making channel island regions, the channel layer 16, 116 is formed on rails 3, 103 and fill layer 30. Layer 16, 116 is then photolithographically patterned into channel islands 161 prior to forming the isolation layer 48, 148. The isolation layer 48, 148 is then formed between and over the channel islands 161. Layer 48, 148 may be planarized to expose the channel islands, or less preferably, a top portion of layer 48, 148 remaining over channel islands may be used as a gate insulating layer

[0078] A gate insulating layer 17 is formed over layer 16. If it is desired to form charge storage transistors 131, then the gate insulating

layer 17 comprises a portion of a charge storage regions, such as an oxide/nitride/oxide film, a dielectric insulated floating gate or conductive nanocrystals embedded in an insulating layer.

[0079] If it is desired to form contacts between different rails, then an additional photoresist mask is formed over layer 17. Layers 17 and 16 are then etched to form vias 60 extending to N+ polysilicon layer 5 in the first rail 3, as shown in Figure 3.

[0080] A second heavily doped semiconductor layer 15 of the first conductivity type is formed over the gate insulating layer 17, as shown in Figure 10D. Preferably, layer 15 is an N+ polysilicon layer if layer 16 is a P- layer, and vice-versa. Layer 15 may be doped in-situ or by ion implantation. If vias 60 are present in layers 16 and 17, then portions of the as-deposited layer 15 contact layer 5 in the first rails 3 through the vias 60.

[0081] A photoresist mask is then formed over the layers 15, 17 and 16 and these layers are etched to form a plurality of second rails 13, as shown in Figure 1. These layers may be etched during one etching step or during plural sequential etching steps, as desired. Alternatively, only layers 15 and 17 may be etched without etching the channel layer 16. For example, since the channel layer 16 is undoped or lightly doped, it may remain as a continuous unpatterned layer between adjacent rails. The active transistor regions are formed where channel layer 16 extends through the openings 49 in the isolation layer 48 to the first rails 3.

[0082] These steps are then repeated to form additional rails 23, 73 over the second rails 13. For example, a second insulating fill layer is formed between the second rails 13. This layer is polished to expose the second rails 13. A second insulating isolation layer 58 is formed over the plurality of second rails 13, as shown in Figure 2. Layer 58 is patterned

to form a plurality of second openings 59 exposing upper portions of adjacent second rails 13. A third lightly doped semiconductor layer 26 is formed over the patterned second isolation layer 58, such that transistor channel portions in layer 26 contact the second rails 13 through the second openings 59, as shown in Figure 2. A gate insulating layer 27 is formed over layer 26. If desired, vias 61 are formed in layers 26 and 27 to allow connection between the second rails 13 and the third rails 23. A third heavily doped semiconductor layer 25 is formed over the gate insulating layer 27. Layers 25, 26 and 27 are patterned to form a plurality of third rails 23, as shown in Figure 1. Alternatively, layer 26 is not patterned and extends between adjacent second rails. A third insulating fill layer is formed between the third rails and polished to expose the third rails 23.

[0083] The method illustrated in Figures 10A-D may be modified depending on what type of array is being fabricated. If the array 1 illustrated in Figure 1 is being fabricated, then the conductivity type of the lightly doped channel layer in each set of rails is reversed. Thus, if layer 16 is p-type, then layer 26 is n-type. Furthermore, each rail except the first rail contains two heavily doped polysilicon layers of opposite conductivity type connected by a metal or a metal silicide layer. The order of these layers in each set of rails is reversed. Thus, in the second rail, a metal or metal silicide layer 19 and a P+ polysilicon layer 18 are deposited onto N+ polysilicon layer 15 in this order. Layers 18, 19, 15, 17 and 16 are patterned together in one or more etching steps to form the second rails 13. In contrast, the P+ polysilicon layer 28 and the metal or metal silicide layer 29 are formed in this order below the N+ polysilicon layer 25 on the gate insulating layer 27. Layers 25, 29, 28, 27 and 26 patterned together in one or more etching steps to form the third rails 23.

[0084] In contrast, if the array 100 illustrated in Figure 8 is being fabricated, then the conductivity type of the lightly doped channel layer in each set of rails is the same. Thus, for example, the channel layer 116, 126 in the second and third rails may comprise an N- amorphous silicon or polysilicon layer. Likewise, each rail contains heavily doped polysilicon layers of the same conductivity type. For example, rails 113 and 123 may contain a single N+ polysilicon layer 125. Alternatively rails 113 and 123 may contain two N+ polysilicon layers separated by a metal or metal silicide layer 129. For example, rail 113 may contain second and fourth N+ polysilicon layers, while rail 123 may contain third and fifth N+ polysilicon layers. Of course, the conductivity types may be reversed in all layers, and p-type semiconductor material may be substituted for n-type semiconductor material. Furthermore, if desired, the silicide layers may be formed by annealing an amorphous silicon/titanium/titanium nitride film as disclosed in U.S. Application Serial No. 09/927,648 filed on August 13, 2002, incorporated herein by reference in its entirety, rather than by reacting a titanium layer with an adjacent polysilicon layer.

[0085] The foregoing description of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed, and modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention. The drawings and description were chosen in order to explain the principles of the invention and its practical application. The drawings are not necessarily to scale and illustrate the device in schematic block format. It is intended that the scope of the invention be defined by the claims appended hereto, and their equivalents.